Introduction to Quantitative Data Analysis

Kaat Smets
mail to: smets@unisi.it

Graduate School in Political Science – University of Siena
Tuesdays 16:00-18:00, First Semester 2010-2011
Computer Lab 2nd Floor

Practical information:

Office hours: Tuesdays 9:30-11:00, Visiting professors room on the third floor (library). During office hours I am pleased to speak with students without an appointment. If you would like to see me at another time, please arrange for an appointment by e-mail.

Website: See the university blackboard at: http://20.unisi.it
The course website contains the course outline, assignments, links to useful websites, etc.

Course description:

Regardless of the level of analysis, political science is ultimately aimed at attempting to answer questions about human behaviour: Why are young people less likely to turn out to vote than older citizens? Does the political ideology of a government influence its welfare state expenditure? What factors explain the severity of civil wars? In seeking answers to such questions, social scientists are interested in the (strength of) the relationship between factors. There are various ways and methods to assess research questions. Quantitative researchers draw on a set of statistical tools to conduct research. This course aims at introducing the basics of such quantitative data analysis.

The course begins with a review of data and data sets. Next, the basic elements of univariate statistics are introduced as we touch upon concepts such as central tendency and dispersion. Measures of association between two variables are discussed as well. Because we want to be able to say something about the degree to which we can be certain our findings say something about the universe, one class will be dedicated solely to significance testing. The remainder of the course will be devoted to different aspects of regression analysis. Regression analysis is the
Introduction to Quantitative Data Analysis

Course outline

foundation for various more complex statistical methods. A good understanding of this analysis technique is, therefore, indispensable.

Because theory and practice go hand in hand, either will be dealt with during this course. Approximately half of the course time will be aimed at putting theory into practice. We will use the statistical software Stata for this empirical part.

Course objectives and learning outcomes:

By the end of the course, participants should be able to:

- Understand key concepts of quantitative methods;
- Interpret results from various types of descriptive analyses;
- Recognize problematic features of data and know how to deal with them;
- Interpret results from a regression analysis

At a practical level, participants will further learn to use Stata.

Course prerequisites:

PhD researchers are expected to have followed the compulsory Stata Crash Course which is taught by Francesco Olmastroni et al. and takes place on October 4-5th 2010. Masters students are strongly encouraged to participate in this two-day introductory course as well.

Set up of the course:

• Meetings:
Each meeting is divided into two parts. The first part is more theoretical and consists of a lecture on the topic that we are discussing in a given week (see the course schedule below for more details). Students are expected to read assigned literature before coming to class. The second part of each class is aimed at putting theory into practice. For this purpose, all classes take place in the computer lab so that we can work through examples using real research questions. Students are recommended to take a usb-stick or an external hard disk with them to be able to store the work that they have done during class.
• Assignments:
Besides practicing in class, students also have to make home work assignments that will be handed out by the end of each meeting. Most – if not all – of these assignments will require the use of a computer. Even when unable to attend class, assignments need to be submitted to me by e-mail by 1pm on the Monday before class. This way I can form an idea of your understanding of the topic and prepare answers to potential questions for the next meeting. Assignments have to include your name, have numbered pages and should be submitted in pdf format using the following scheme to name your document: Firstname Lastname - assignment xx.pdf. If you foresee any problems handing in your assignment you should contact me well before the deadline. Late assignments will not be accepted and you will have to do an extra assignment in case you fail to meet requirements.

• Examination:
Your grade will be determined by a set of assignments (40% of the total grade), and a final exam (60% of the total grade).

Literature:

Classes will largely follow the contents of the following books:

- Kohler, Ulrich and Frauke Kreuter. 2009. Data Analysis Using Stata. 2nd vol. College Station, TX: Stata Press.


Each of these books is available in the university library, although you will have to go to the Faculty of Economics for the Kohler & Kreuter book. They are also available for less than £20 respectively on e.g. amazon.co.uk. I may occasionally hand out additional readings during the course.
Course schedule:

- **Week 1 (Tuesday, October 12th 2010)**

  What is a data set? Where do we find them? What different kinds of data sets do political scientists use? What different kind of variables do we work with?

  *To read:*
  - *Lewis-Beck (1995), chapter 1 & 2*
  - *Kohler and Kreuter, chapter 1 & 2*

- **Week 2 (Tuesday, October 19th 2010)**

  Describing data: central tendency and dispersion; Measures of association

  *To read:*
  - *Lewis-Beck (1995), chapter 3 & 4*
  - *Kohler and Kreuter, chapter 7*

- **Week 3 (Tuesday, October 26th 2010)**

  Significance testing: central limit theorem, type I & II errors

  *To read:*
  - *Lewis-Beck (1995), chapter 5*

- **Week 4 (Tuesday, November 2nd 2010)**

  What have we learned so far?
• Week 5 (Tuesday, November 9th 2010)

Introducing (bivariate) regression analysis

To read:
- Lewis-Beck (1995), chapter 6
- Lewis-Beck (1989), chapter 1
- Kohler & Kreuter, section 8.1

• Week 6 (Tuesday, November 16th 2010)

Assumptions of regression analysis

To read:
- Lewis-Beck (1989), chapter 2

• No class on Tuesday, November 23rd 2010

• Week 7 (Tuesday, November 30th 2010)

Multivariate regression analysis

To read:
- Lewis-Beck (1995), chapter 7 up to p. 60
- Lewis-Beck (1989), chapter 3 up to p. 54
- Kohler & Kreuter, section 8.2

• Week 8 (Tuesday, December 7th 2010)

Regression diagnostics

To read:
- Kohler & Kreuter, section 8.3
• No classes from December 14th 2010 to January 4th 2011

• Week 9 (Tuesday, January 11th 2011)

Dummy variables, interaction effects, transformations

To read:
- Lewis-Beck (1995), chapter 7 from p. 60 onwards
- Lewis-Beck (1989), chapter 3 from p. 54 onwards
- Kohler & Kreuter, section 8.4

• Week 10 (Tuesday, January 18th 2011)

Wrap up session

• Week 11 (Tuesday, January 25th 2011 – tbc)

Final exam

A last note:

Please note that this course outline may be subject to change. I will make sure to inform you of any such changes. Up-to-date information on scheduled meetings, class room changes, the course outline, etc. can also be found on the course website.